信息与熵力-生物细胞的物理描述 ——化学反应动力学和信息论

钱 紘

Department of Applied Mathematics, University of Washington, WA 98195-3925, USA

E-mail: hqian@u.washington.edu

收稿日期: 2016-09-06; 接受日期: 2016-10-22; 网络版发表日期: 2016-12-08

摘要 化学动力学和信息论, 作为物理学思维与方法的发展, 在描述生物细胞个体的行为、逻辑及功能中有着极其重要的地位. 基于最近随机化学动力学与吉布斯化学热力学统一理论的发展, 本文尝试着将化学动力学与信息论在同一个随机动力学的框架内讨论, 提出了信息为熵力之假设.

关键词 细胞, 生物物理, 统计力学, 非平衡态, 熵力, 个体, 种群, 涌现现象

自香农(Shannon)的信息论问世以来, 其在工程与科学各领域一直有着显赫的地位^[1]. 近年来, 信息论在生物学中的应用也是与日剧增^[2]. 作为定量刻画通讯中编码传递的理论, 香农的学说无可置疑; 但是作为定量描述生命过程的一个视角, 它与物理、化学、动力学、统计学的根本区别又是什么呢? 要准确地回答这个命题, 首先问一个看似平凡但实际上更基本的疑问: “信息”(information)到底是什么? 它不是物质, 那它是某一种能量吗? 物理学给予了人们对物质和能量深刻的认识: 力学告诉我们万物是怎么样和为什么运动的; 化学告诉我们物质是如何可以由简单组成复杂; 动力学(dynamics), 作为自牛顿以来应用数学的核心, 用充分甚至于是必要的因果关系来描述世界的千变万化; 而统计学则教我们认识到个体与种群之间辩证的关系, 以及自然界的规律也可从严格的数学中发现: 正态分布在周围到处都能看到, 它却可以在

统计学里完全用数学演绎导出. 基于以上这些粗线条的框架, 发现信息论, 作为通讯工程之外的一个普适的理论, 与力学有着共同的追求——两者都讲因果关系; 它与概率论和统计学又运用着同样的数学逻辑与方法. 这也许是人们一直在试图用信息论来顶替统计力学的原因^[3].

事实上, 广义的信息论似乎与另外一支新兴的学科, 随机动力学(stochastic dynamics)有着异曲同工之处. 后者讲究的是以机制出发的(mechanism-based)随机数学模型做为科学的语言来刻画世界. 它与以数据出发的(data-based)随机数学模型和以机制出发的确定性数学模型都有区别; 前者为统计学的宗旨, 后者为牛顿以降古典的应用数学之精华. 无怪乎H. Haken在他的协同学(synergetics)中不断地讲非线性随机动力学和信息论^[4,5]. 用随机动力学的数学方法来描述世界起源于赌博^[6]、股票^[7]和布朗运动^[8]. 在此之后, 许

引用格式: 钱紘. 信息与熵力-生物细胞的物理描述——化学反应动力学和信息论. 中国科学: 生命科学, 2017, 47: 257-261

Qian H. Information and entropic force: physical description of biological cells, chemical reaction kinetics, and information theory. Sci Sin Vitae, 2017, 47: 257-261, doi: [10.1360/N052016-00242](https://doi.org/10.1360/N052016-00242)

多重重要概念的提出都有深厚的物理化学背景和概率统计内涵: 吉布斯(Gibbs)^[9]告诉我们平稳概率分布的对数改号在热力学平衡态是一种能量, 既吉布斯能. Kirkwood^[10]告诉我们自由能事实上是平衡态里的平均力的势函数. 昂萨格(Onsager)^[11]则告诉我们新的形式的力会在非平衡态系统中以“热力学力”的形式涌现出来, 而与此涌现力(emergent force)对应的恰恰是刻画运输过程的“热力学流”. 那么称此热力学力为其对应的热力学流的原因(cause)是否合理呢? 能否甚至说前者是后者的“机制”呢? 这些都是面对多尺度复杂系统时我们不可回避的深刻问题. 随机动力学的一个成功的例子是高分子动力学(polymer dynamics)^[12]. 而橡胶弹性(rubber elasticity)的合理解释正是熵力这个重要概念的应用^[13].

1 信息为熵力之假设

让我们回到信息是什么的问题. 它是不是一种能量呢? 香农信息论中最重要的概念是熵. 热力学里的熵, 或者说负熵乘以温度, 的确是自由能的一部份; 化学里的熵可以当作一种力来做功^[13], 熵弹簧之刚性随温度增加而增加. 所以我认为香农熵与热力学熵的最大不同是后者, 通过绝对温度的测量及玻尔兹曼常数, 已经与牛顿力学统一在一起了. 而前者还只是一个无量纲的概念. 它的完全定量, 以及测量, 有待于多尺度复杂系统中的“力”、“位移”和“功”等一系列概念的发展^[14].

如果称前文的论述为我们的“信息为熵力之假设”, 那么用随机化学动力学来描述细胞内的生物化学过程则可以很好地验证此假设. 更有甚者, 对细胞生物学的基因调控与信号传导的描述已经大量地运用着信息论的语言和思想^[15-17]. 而随机化学动力学^[18-20]恰好也在最近得到了与吉布斯化学热力学的统一^[21-24], 而且被推广到远离平衡态的定态^[25-27], 成为了一套相当完美的跨越介观到宏观的数学理论.

一个随机动力学系统经过很长时间后会达到一个定态. 这里所说的定态是一个统计概念. 一个确定性动力系统的定态是相空间内的一个点. 多稳系统则有多个可能的定态, 但每个特定系统的定态是由其初条件决定的. 很多确定性动力系统长时间的形状根本不是定态, 而是振荡. 一个随机动力学系统的定态是

一种具有平稳性的随机振荡, 是一种“state of motion”. 描述随机定态不再是相空间内的一些点, 或者几条轨道. 而是整个相空间上的一个概率分布, 通常被称为不变分布. 很多时候这个分布是唯一的, 或者有物理意义的不变分布通常是唯一的.

不变的概率分布蕴含着一个力, 直观上这很容易理解. 一个动力学系统总是更经常地从小概率的地方向大概率的地方移动, 而一件总是发生的事情就可以用一个“力”来表述, 进一步来定量地刻画, 这就是熵力. 对于一个信奉还原论的学者来说, 为什么这里概率大那里概率小的真正原因存在于系统的动力学中. 熵力实际上是对周而复始的复杂运动的一个粗粒化的概括和总结, 而不是其原因. 但是这样定量的描述, 如果系统的运动会重复, 是有预测之功能的. 所以熵力很像中国人所说的“成者为王, 败者为寇”. 西方人称之为the end justifies the means.

2 香农熵和自信息

熵首先是个数学概念. 它存在于每一个概率统计模型中. 从数学的角度来说, 任何一个离散随机模型都是建立在一个概率空间 $(\Omega, \mathcal{F}, p)$ 上的. 在技术层面, 香农熵是定义在此概率空间之上的一个数^[1]:

$$H[\Omega] = - \sum_{z \in \Omega} p_z \ln p_z,$$

而M. Tribus则进一步引进了一个随机自信息(self-information)的概念, 它是个定义在同一个概率空间上的随机变量, $-\ln p_z$, 其数学期望值为香农熵^[28,29]. 自信息具有深刻的概率统计内涵: 它告诉我们概率空间是较随机变量更为根本的东西. 一个随机变量 $X(\omega)$ 只是对其背后的概率空间的一个观测量, 它很可能是片面的(biased). 而自信息 $-\ln p_X$ 则是在最大程度上由一个片面的观测量 $X(\omega)$ 还原到概率空间自己. 假设 $\Omega \rightarrow X(\omega)$ 的对应是一一对应的, 则:

$$\begin{aligned} H_X &= - \sum_{X(\omega)} \Pr\{X(\omega)\} \ln \Pr\{X(\omega)\} \\ &= - \sum_{z \in \Omega} p_z \ln p_z = H[\Omega], \end{aligned}$$

否则的话 $H_X \leq H(\Omega)$. 概率空间的香农熵具有一种极大性. 事实上同一个概率空间上的两个不同的“全息”观测量有相同的香农熵. 观测量背后存在着一个抽象的

数学空间的思想起源于量子力学^[30]. 方差、条件方差之期望、协方差等统计量离数据很近, 信息论里的熵、条件熵、互信息(mutual information)则离概率空间很近.

3 相对熵与动力学

如果说自信息和熵是对于一个周而复始的, 随机振荡的动力系统的定态之概率不变分布 π 的刻画, 那么相对熵:

$$H[p(t) \parallel \pi] = \sum_{z \in \Omega} p_z(t) \ln \frac{p_z(t)}{\pi_z},$$

则与这个系统的瞬时状态的变化, 也就是其动力学, 有着千丝万缕的联系, 这里 $p(t)$ 是一个随机动力学系统在时间 t 时相空间上的统计分布. 一个动力系统可以是离散时间的, 也可以是连续时间的. 前者的例子是一维单峰映射如 $x_{n+1} = \alpha x_n(1-x_n)$; 这个数学模型常常被用来描述昆虫野生种群上一代(x_n)与下一代(x_{n+1})数量间的关系^[31]. 后者的例子是微分方程; 描述酶催化反应的米氏(Michaelis-Menten)方程属于这一类. 可以证明 $H[p_z(t) \parallel \pi]$ 首先是非负的. 而且随着时间 t 它不会增长. 更准确地说, 在完全确定的一对一的动力系统中它是不变的(上面的单峰映射就不是一对一的: $X_n=0.8$ 和 0.2 会得到同样的 $x_{n+1}=0.16\alpha$, 它是个二对一的映射). 但对含有不确定性的动力系统它一定是递减的^[32]. 多对一动力系统的过去有不确定性; 一对多动力系统的未来有随机性. 只有一对一的动力系统才有清楚的“来龙去脉”. 热力学统计物理中的自由能正是相对熵此数学量的体现^[21,33].

如果一个随机动力学系统在某一极限下趋于一个确定性的动力系统, 那么前者的分布 $p_z(t)$ 趋于一个确定性运动 $\bar{z}(t)$. 这时候 $p_z(t)$ 的香农熵成为零, 而相对熵 $H[p_z(t) \parallel \pi]$ 成为:

$$-\sum_{z \in \Omega} p_z(t) \ln \pi_z \rightarrow -\ln \pi_{\bar{z}(t)},$$

相对熵随时间递减的事实转化成 $-\ln p_z$ 为确定性动力学的能量函数^[23,24].

4 作为介观化学动力学系统的单细胞个体

生物过程, 不同于物理过程, 是有功能的^[34], 这一

论点在近年来得到极大的认可. 可是我们需要意识到分子只是运动着, 复杂系统的运动会呈现出意想不到的规律. 把物理化学的运动和规律与生物学的现有知识、逻辑, 以及叙述(narrative)结合起来, 说来说去是人们对以生物化学为基础的生命世界的一种诠释. 细胞生物学对信号传导的描述充满了叙述性的语言, 这也许有助于对复杂系统的知识的整合, 但未必离真实更近. 动力学的最基本的思想是在“相空间”里讨论运动, 而相空间里的一个点是由所有的参与者(players)的共同状态确定的.

蛋白分子内的原子运动非常复杂. 两个血红蛋白(hemoglobin)分子内相对应的原子有着完全不同的运动. 但是这个事实并没有阻止化学家把具有相同化学分子式的分子, 尤其是在同样的温度和压强之下, 当作“同类”. 同一个血红蛋白实际上还有T与R构象之区别. 一个蛋白分子不同构象间的转换是蛋白质物理化学的重要课题之一. 如果把同样的化学动力学的思想运用到单细胞上, 那么具有相同基因组的细胞个体, 在同样的培养液中, 应当作同类. 细胞个体的不同表型, 以及不同表型间的转换则应该是细胞物理化学的重要课题. 近年来细胞生物学的热点恰恰反映了这一看法的成立. “同类细胞”的不同表型, 在适当的生物组织(biological tissue)内, 当然会导致其不同的生物功能: 表皮细胞与干细胞的区别以及它们之间的转换正可以通过随机化学动力学^[18-20]的方法来研究.

5 结语

物理学是对客观世界普适又深刻的描述与理解; 它追求用一个统一的数学框架来定量地刻画事件的发生及过程. 生物过程看上去和物理学研究的对象非常不一样: 它们为“多体”(many-body)复杂现象^[35]. 虽然复杂系统的一切现象最终是由其所有的粒子或个体的运动决定的, 但常常作为一种涌现现象(emergent phenomenon)它可以用更简明和准确的抽象语言来描述. 概率统计正是一门在这方面极为有用的数学方法; 而信息论则把熵的概念提到了一个“主宰一切”的高度. 以随机动力学为起点, 本文说明平稳态的概率分布蕴含着一个“热力学力”的概念; 香农熵和自信息恰恰反映了这种熵力. 在涌现现象的层面上, 熵力就像物理学其他分支里的力一样, 有待人们去发现或建构^[36].

参考文献

- 1 Shannon C E, Weaver W. The Mathematical Theory of Communications. Champaign: The University of Illinois Press, 1971
- 2 Cronin L, Walker S I. Beyond prebiotic chemistry. [Science](#), 2016, 352: 1174–1175
- 3 Jaynes E T. Probability Theory: The Logic of Science. Cambridge: Cambridge University Press, 2003
- 4 Haken H. Synergetics—An Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology. New York: Springer-Verlag, 1983
- 5 Haken H. Information and Self-Organization: A Macroscopic Approach to Complex Systems. New York: Springer, 2010
- 6 Montroll E W, Shlesinger M F. On the wonderful world of random walks. In: Lebowitz J L, Montroll E W, eds. Studies in Statistical Mechanics, vol. XI (Nonequilibrium Phenomena II: From Stochastics to Hydrodynamics) Amsterdam: North-Holland, 1984. 5–121
- 7 Bachelier, L. Théorie la spéculation. Annales Scientifiques de l'École Normale Supérieure, 1900, 3: 21–86
- 8 Einstein A. Investigations on the Theory of the Brownian Movement. New York: Dover, 1956
- 9 Gibbs J W. The Scientific Papers of J. Willard Gibbs, vol. I and II. London: Longmans Press, 1906
- 10 Kirkwood J G. Statistical mechanics of fluid mixtures. J Chem Phys, 1935, 3: 300–313
- 11 Onsager L. Reciprocal relations in irreversible processes. I. Phys Rev, 1931, 37: 405–426
- 12 Doi M, Edwards S F. The Theory of Polymer Dynamics. Oxford: Clarendon Press, 1988
- 13 Hill T L. Statistical Mechanics: Principles and Selected Applications. New York: Dover, 1987
- 14 Ma Y A, Qian H. A thermodynamic theory of ecology: Helmholtz theorem for Lotka-Volterra equation, extended conservation law, and stochastic predator-prey dynamics. Proc Roy Soc A, 2015, 471: 20150456
- 15 Thomas P J. Every bit counts. [Science](#), 2011, 334: 321–322
- 16 Qian H, Roy S. An information theoretical analysis of kinase activated phosphorylation dephosphorylation cycle. [IEEE Transon Nanobiosci](#), 2012, 11: 289–295
- 17 Bialek, W. Biophysics: Searching for Principles. Princeton: Princeton University Press, 2013
- 18 Beard DA, Qian H. (2008) Chemical Biophysics: Quantitative Analysis of Cellular Systems. Cambridge: Cambridge University Press, 2008
- 19 Érdi P, Lente G. Stochastic Chemical Kinetics: Theory and (Mostly) Systems Biological Applications. New York: Springer-Verlag, 2014
- 20 Anderson D F, Kurtz T G. Stochastic Analysis of Biochemical Systems. New York: Springer, 2015
- 21 Ge H, Qian H. Physical origins of entropy production, free energy dissipation, and their mathematical representations. [Phys Rev E](#), 2010, 81: 051133 [0911.3984](#)
- 22 Qian H, Kjelstrup S, Kolomeisky A B, et al. Entropy production in mesoscopic stochastic thermodynamics: nonequilibrium kinetic cycles driven by chemical potentials, temperatures, and mechanical forces. [J Phys-Condens Matter](#), 2016, 28: 153004 [1601.04018](#)
- 23 Ge H, Qian H. Mesoscopic kinetic basis of macroscopic chemical thermodynamics: a mathematical theory. 2016, arXiv: 1601.03159
- 24 Ge H, Qian H. Mathematical formalism of nonequilibrium thermodynamics for nonlinear chemical reaction systems with general rate law. 2016, arXiv: 1604.07115
- 25 Zhang X J, Qian H, Qian M. Stochastic theory of nonequilibrium steady states and its applications (Part I). Phys Rep, 2012, 510: 1–86
- 26 Ge H, Qian M, Qian H. Stochastic theory of nonequilibrium steady states (Part II): applications in chemical biophysics. Phys Rep, 2012, 510: 87–118
- 27 葛颢. 非平衡态统计物理的随机数学理论. 数学进展, 2014, 43: 161–174
- 28 Cover T M, Thomas J A. Elements of Information Theory. New York: Wiley, 1991
- 29 Qian H. Mesoscopic nonequilibrium thermodynamics of single macromolecules and dynamic entropy-energy compensation. Phys Rev E, 2002, 65: 016102
- 30 von Neumann J. Mathematical Foundations of Quantum Mechanics. Princeton: Princeton University Press, 1955
- 31 May R M. Simple mathematical models with very complicated dynamics. Nature, 1976, 261: 459–467
- 32 Voigt J. Stochastic operators, information, and entropy. Comm Math Phys, 1981, 81: 31–38
- 33 Mackey M C. The dynamic origin of increasing entropy. Rev Mod Phys, 1989, 61: 981–1015
- 34 Hartwell L H, Hopfield J J, Leibler S, et al. From molecular to modular cell biology. [Nature](#), 1999, 402: C47–C52
- 35 Hopfield J J. Physics, computation, and why biology looks so different. [J Theor Biol](#), 1994, 171: 53–60
- 36 Anderson P W. More is different-Broken symmetry and the nature of the hierarchical structure of science. [Science](#), 1972, 177: 393–396

Information and entropic force: physical description of biological cells, chemical reaction kinetics, and information theory

QIAN Hong

Department of Applied Mathematics University of Washington Seattle, WA 98195-3925, USA

Chemical reaction kinetics and information theory, as continuous developments of the methodologies and thinkings originating in physics, have played prominent roles in describing the behaviors, logics, and functions of individual biological cells. Based on a recently developed unifying theory of mesoscopic chemical kinetics and Gibbsian macroscopic chemical thermodynamics, this paper attempts to discuss both chemical kinetics and information theory in the single framework of nonlinear, stochastic dynamical systems, and proposes the hypothesis that information is a form of entropy force.

cell, biophysics, statistical mechanics, nonequilibrium, entropic force, individual, population, emergent phenomenon

doi: [10.1360/N052016-00242](https://doi.org/10.1360/N052016-00242)